No. of Printed Pages : **8**

A10.5-R5 : MACHINE LEARNING USING PYTHON

DURATION : 03 Hours **MAXIMUM MARKS : 100**

OMR Sheet No. :						

Roll No. : | | | | | | | **Answer Sheet No. :** | | | | | | |

Name of Candidate : _____ **; Signature of Candidate :** _____

INSTRUCTIONS FOR CANDIDATES :

- Carefully read the instructions given on Question Paper, OMR Sheet and Answer Sheet.
- Question Paper is in English language. Candidate has to answer in English language only.
- There are **TWO PARTS** in this Module/Paper. **PART ONE** contains **FOUR** questions and **PART TWO** contains **FIVE** questions.
- **PART ONE** is Objective type and carries **40** Marks. **PART TWO** is Subjective type and carries **60** Marks.
- **PART ONE** is to be answered in the **OMR ANSWER SHEET** only, supplied with the question paper, as per the instructions contained therein. **PART ONE** is **NOT** to be answered in the answer book for **PART TWO**.
- Maximum time allotted for **PART ONE** is **ONE HOUR**. Answer book for **PART TWO** will be supplied at the table when the Answer Sheet for **PART ONE** is returned. However, Candidates who complete **PART ONE** earlier than one hour, can collect the answer book for **PART TWO** immediately after handing over the Answer Sheet for **PART ONE** to the Invigilator.
- **Candidate cannot leave the examination hall/room without signing on the attendance sheet and handing over his/her Answer Sheet to the invigilator. Failing in doing so, will amount to disqualification of Candidate in this Module/Paper.**
- After receiving the instruction to open the booklet and before answering the questions, the candidate should ensure that the Question Booklet is complete in all respects.

DO NOT OPEN THE QUESTION BOOKLET UNTIL YOU ARE TOLD TO DO SO.

PART ONE

(Answer all questions; each question carries ONE mark)

1. Each question below gives a multiple choice of answers. Choose the most appropriate one and enter in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

1.1 A feature F1 can take a certain value : A, B, C, D, E, or F, which represents the grades of students from a college. Which of the following statement is **true** in the following case ?

(A) Feature F1 is an example of a nominal variable.

(B) Feature F1 is an example of an ordinal variable.

(C) It doesn't belong to any of the above categories.

(D) Both (A) and (B)

1.2 Which of the following is an example of a deterministic algorithm ?

(A) PCA

(B) K-Means clustering

(C) KNN

(D) None of the above

1.3 Below are the 8 actual target variable values in the training file.

[0,0,0,1,1,1,1,1] What is the entropy of the target variable ?

(A) $-(5/8 \log(5/8) + 3/8 \log(3/8))$

(B) $5/8 \log(5/8) + 3/8 \log(3/8)$

(C) $3/8 \log(5/8) + 5/8 \log(3/8)$

(D) $5/8 \log(3/8) - 3/8 \log(5/8)$

1.4 Let's say you are using activation function X in hidden layers of a neural network. At a particular neuron for any given input, you get the output as "-0.0001". Which of the following activation function could X represent ?

(A) ReLU

(B) tanh

(C) SIGMOID

(D) None of the above

1.5 Which of the following is/are one of the important step(s) to pre-process the text in NLP-based projects ?

(1) Stemming

(2) Stop word removal

(3) Object Standardization

(A) (1) and (2)

(B) (1) and (3)

(C) (2) and (3)

(D) (1), (2) and (3)

1.6 What is morphological segmentation ?

(A) Separate words into individual morphenes

(B) Does disclosure analysis

(C) Is an extension of postpositional logics

(D) Does sentiment analysis

1.7 Which of the following is used to mapping sentence plan into sentence structure ?

(A) Text planning

(B) Sentence planning

(C) Text Realization

(D) None of the above

1.8 In Model based learning methods, an iterative process takes place on the ML models that are built based on various model parameters, called _____.

(A) mini-batches

(B) optimizedparameters

(C) hyperparameters

(D) superparameters

1.9 The two key ideas of deep learning for computer vision :

(A) Convolutional neural networks and backpropagation

(B) Deep neural networks and kernel functions

(C) Support Vector Machines and loss functions

(D) None of the above

1.10 Example of reinforcement learning in machine learning :

(A) Parking can be achieved by learning automatic parking policies

(B) Enjoying ice cream sitting in automatic car

(C) Machines are super machines, one can do every task

(D) None of the above

2. Each statement below is either TRUE or FALSE. Choose the most appropriate one and enter your choice in the "OMR" answer sheet supplied with the question paper, following the instructions therein. (1x10)

2.1 A Pearson correlation between two variables is zero; still, their values can be related to each other.

2.2 LogLoss evaluation metric can have negative values.

2.3 Deep learning is a digital framework for learning representations from data.

2.4 Loss function is same as objective function.

2.5 Learning means finding a set of values for the weights of all layers in a network, such that the network will correctly map example inputs to their associated targets.

2.6 Pattern classification involves unsupervised learning while grouping is a supervised one.

2.7 Gradient at a given layer is the product of all the gradients at the previous layers.

2.8 The list.pop ([i]) removes the item at the given position in the list.

2.9 Python Literals is used to define the data that is given in a variable or constant.

2.10 If K=N, then K-fold cross-validation is called Leave one out cross validation, where N is the number of observations.

3. **Match words and phrases in column X with the closest related meaning/word(s)/phrase(s) in column Y. Enter your selection in the "OMR" answer sheet supplied with the question paper, following the instructions therein.** **(1x10)**

X		Y	
3.1	steps of NLP	**A.**	SVM
3.2	Kernel	**B.**	5
3.3	CNN	**C.**	Non-linearly separable data
3.4	Backpropagation algorithm	**D.**	Image processing
3.5	The range of the tanh function is	**E.**	Type I error
3.6	Incorrect rejection of a true null hypothesis	**F.**	[-1,1]
3.7	Python library used for working with arrays	**G.**	[0,1]
3.8	the proportion of actual positive cases which are correctly identified.	**H.**	Specificity
3.9	The range of the RELU function is	**I.**	Sensitivity
3.10	The proportion of actual negative cases which are correctly identified	**J.**	Type II error
		K.	[0,infinity]
		L.	NumPy
		M.	Sklearn

4. Each statement below has a blank space to fit one of the word(s) or phrase(s) in the list below. Choose the most appropriate option, enter your choice in the "OMR" answer sheet supplied with the question paper, following instructions therein. (1x10)

A.	Training	**B.**	Shallow learning	**C.**	Key Value pair
D.	Tensorflow	**E.**	Stemming	**F.**	Seaborn
G.	dist()	**H.**	Pandas	**I.**	Overfitted model
J.	Voice recognition	**K.**	Range()	**L.**	Regression
M.	Super fitted model				

4.1 In linguistic morphology _____ is the process for reducing inflected words to their root form.

4.2 A model that learns the training dataset too well, performing well on the training dataset but does not perform well on a testing dataset _____.

4.3 Machine learning tend to focus on learning only one or two layers of representations of the data is _____.

4.4 Example of an unsupervised feature map _____.

4.5 Process of improving accuracy of a neural net is _____.

4.6 Python Dictionary is used to store the data in a _____ format.

4.7 The _____ is a built-in function that returns a range object that consists of series of integer numbers, which we can iterate using a for loop.

4.8 Python _____ library is a widely popular for data visualization.

4.9 Python _____ library is a widely popular for data manipulation and analysis.

4.10 _____ is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome.

PART TWO

(Answer any FOUR questions)

5. (a) What are Support Vectors in SVM ? Explain.

(b) What are the different Kernels in SVM ? Explain.

(c) How do you choose which machine learning algorithm to use (elaborate all the steps) ? Support your answer with a suitable example. **(4+4+7)**

6. (a) What is PCA ? What circumstances do you use it in ?

(b) What is the difference between Gini Impurity and Entropy in a Decision Tree ?

(c) How would you screen for outliers, and what should you do if you find one ? **(5+5+5)**

7. (a) Running a binary classification tree algorithm is quite easy. But do you know how the tree decides on which variable to split at the root node and its succeeding child nodes ?

(b) How can you select K for K-means Clustering ?

(c) What exactly is a Random Forest ? What is the mechanism ? **(5+5+5)**

8. (a) How do you define a neural network ? Explain all its elements.

(b) What do Cost Functions and Loss Functions mean ? Describe their main differences.

(c) When should one use an MLP, CNN, or RNN ? Give example. **(5+5+5)**

9. (a) Name a few libraries in Python used for Data Analysis and Scientific Computations. Which library would you prefer for plotting in Python language ?

(b) What is the main difference between a Pandas series and a single-column Data Frame in Python ?

(c) What is the difference between tuples and lists in Python ?

(d) What is aggregation? Briefly discuss different methods available in Python to perform aggregations on data. **(3+3+4+5)**

- o O o -

SPACE FOR ROUGH WORK

SPACE FOR ROUGH WORK